雨生红球藻低覆盖度基因组草图分析

陈军[①②]，郑华军[③*]，刘亚铭[①②]，赵国屏[③]，秦松[①*]
① 中国科学院烟台海岸带研究所海岸带生物学与生物资源利用重点实验室, 山东烟台 264003
② 中国科学院大学，北京 100049
③ 上海人类基因组研究中心，上海 201203

摘要：开展雨生红球藻基因组测序研究，对于解读绿藻起源与进化及生物逆境胁迫响应机理，及推动雨生红球藻产业发展都具有重要意义。利用 Illumina Hiseq 2500 对雨生红球藻(*Haematococcus pluvialis*)进行高通量测序，获得低覆盖度全基因组草图。通过计算 k-mer 分布预测该基因组草图大小约为 547 Mbp，GC 含量为 59.2 %，为纯合或单倍体。共得到 11,059 个预测基因，平均基因长度为 1,711 bp，平均 CDS 长度为 681 bp；平均每个基因包含 3.2 个外显子，外显子平均长度为 353 bp。代谢通路分析表明，具有完整的糖酵解、三羧酸循环、磷酸戊糖途径、嘌呤和嘧啶合成等基本代谢通路。
关键词：雨生红球藻；基因组测序；基因预测；功能注释

Title: The analysis of the low coverage *Haematococcus pluvialis* draft genome

Chen Jun[①②], Zheng Hua-jun[③*], Liu Ya-ming[①②], Zhao Guo-ping[③], Qin Song[①*]
①Key Laboratory of Coastal Biology and Bioresource Utilization, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003
②University of Chinese Academy of Sciences, Beijing 101418
③Chinese National Human Genome Center at Shanghai, Shanghai 201203

Keywords: *Haematococcus pluvialis*; genome sequencing; gene prediction；gene functional annotation
Abstract: The investigation on the genomic study of *Haematococcus pluvialis* would be significant to explore the origin and evolution of green algae and the stress responses in *Haematococcus pluvialis,* and promote the development of *Haematococcus pluvialis* industry. The low-coverage draft genome of *Haematococcus pluvialis* was constructed by the Illumina Hiseq 2500 platform. The predicted genome size was approximately 547 Mbp, with the GC content of 59.2% by calculating K-mer distribution. The draft genome contained 11,053 predicted protein-coding genes and the average gene size and CDS were 1,711 and 681 bp; every gene contained 3.2 exons and the size of exon was 353 bp in average. The analysis of metabolic pathway indicated that the low-coverage genome contained whole glycolysis, tricarboxylic acid cycle, phosphopentose, purine and pyrimidine synthesis and other basic metabolism pathway.

第一作者：陈军，E-mail: junchen@yic.ac.cn;
通信作者：郑华军，E-mail: zhenghj@chgc.sh.cn;秦松，E-mail：sqin@yic.ac.cn
资助项目情况：国家重点研发计划(2016YFF0202304)；国家自然科学基金面上项目(41376139)；烟台市科技计划(2016JHZB007)
Foundation: National key research and development program-China (2016YFF0202304); National Natural Science Foundation of China (41376139); Science and Technology Program of Yantai City (2016JHZB007).

雨生红球藻（*Haematococcus pluvialis* Flotow 1844）是一种单细胞真核微藻，隶属绿藻门（Chlorophyta），绿藻纲（Chlorophyceae），团藻目（Volvocales），红球藻科（Haematococcaceae），红球藻属（*Haematococcus*），分布在各种小水体和潮湿土壤[1]。雨生红球藻具有复杂的生活周期，分为游动细胞、不动细胞、动孢子和不动孢子期[2,3]。在特定环境条件下，如高光、高盐、缺氮等，雨生红球藻从游动细胞转变为不动孢子，并累积虾青素，约占其细胞干重的1%-4%，被认为是虾青素的最好来源[4,5]。天然虾青素具有超强的抗氧化活性，有"超级维生素E"之美誉，生产能力远不能满足市场需求，国际市场价格高达10,000美元/kg，大约是人工合成虾青素的200倍，广泛应用在食品保健品、化妆品、饵料饲料等领域[5,6]。

为解读雨生红球藻合成虾青素的分子机制，摸清虾青素代谢机理，指导选育优良品种，建立雨生红球藻合成虾青素代谢模型，近年来，雨生红球藻分子生物学研究取得了很大进步。虾青素在雨生红球藻中的代谢途径已经基本清楚[7-9]。伴随着高通量测序技术的发展，雨生红球藻转录组学[10-14]、蛋白组学[15-17]和代谢组学[18, 19]研究也不断增多。Chen 等[13]基于虾青素在雨生红球藻细胞内以虾青素酯的形式存在的特点，利用转录组学和代谢组学的手段，阐释了雨生红球藻虾青素和脂肪酸合成的协同调控分子机理，表明虾青素酯化过程对于虾青素的形成和积累具有推动作用。Gwak 等[14]利用转录组学和脂组学的手段，分析了雨生红球藻绿色游动细胞和红色不动孢子期细胞虾青素合成和脂肪酸代谢规律，重点阐述了高光胁迫条件下胞囊形成过程中，雨生红球藻虾青素合成和脂肪酸代谢协同调控机制。

然而，外界逆境如何精准调控雨生红球藻合成虾青素？雨生红球藻合成虾青素代谢通路中编码关键限速酶基因上游有哪些顺式作用元件？这些顺式作用元件如何与转录因子相互作用？雨生红球藻积累虾青素是细胞响应外界逆境，清除活性氧自由基时产生的副产物还是细胞产生虾青素的目的是为了清除自由基？雨生红球藻生物合成虾青素仍有诸多问题需要解决[20]，而通过传统 PCR、转录组学等方法获得的序列，信息往往不全面，因此，很有必要对雨生红球藻进行全基因组测序，以便于对其遗传信息进行更加系统的研究。本研究利用第二代高通量测序仪 Illumina Hiseq 2500 对雨生红球藻进行低覆盖度全基因组测序，预测了雨生红球藻基因组大小等基因组基本特征,为绘制高质量雨生红球藻基因组精细图作了准备。

1. 材料和方法

1.1 实验材料及培养条件

本研究使用的雨生红球藻（*Haematococcus pluvialis*）购自英国 CCAP 藻种库，现保存于中国科学院烟台海岸带研究所海岸带生物学与生物资源利用重点实验室。雨生红球藻接种于 BBM 培养基，于 25℃光照条件下静止培养，光照强度为 40 μmol/(m^2·s)，光/暗周期为 12 h/12 h。

1.2 Illumina 基因组文库构建和测序

取处于对数生长期的雨生红球藻细胞，利用天根新型植物基因组 DNA 提取试剂盒（DP320-02，北京）提取基因组，然后使用 1 %的琼脂糖凝胶电泳进行检测，并利用超微量分光光度计测定 DNA 浓度和纯度（A260/A280）。获得高质量的基因组后，构建 Pair-end 文库。利用 Covaris S2（Covaris，美国）仪器将 DNA 打断至 400 bp 大小的片段，利用 TruSeq^{TM} DNA Sample Prep Kit – Set A (Illumina，美国) 制备文库，最后切胶回收 350- 450 bp 的片段。取 10 ng 构建好的文库 DNA，用 TruSeq PE Cluster Kit (Illumina，美国) 在 cBot 中进行 cluster generation，然后在 Illumina Hiseq^{TM} 2500 中进行双向测序。

1.3 数据处理

通过高通量测序得到的原始数据使用软件 FASTX-Toolkit（http://hannonlab.cshl.edu/fastx_toolkit/）得到 clean 序列，具体步骤如下：(1)使用 fastx_clipper 去除 reads 中的接头序列；(2)使用 fastq_quality_filter 从 read 的 3’端到 5’端方向开始去除 N 至第一个不是 N 的碱基；(3)使用 fastq_quality_filter 从 read 的 3’端到 5’端方向开始去除连续出现的低质量碱基(q < 5)，当去除低质量的 clean reads 长度低于 50bp 时删除 read 本身及其配对的序列；(4)使用本地脚本配对 pair end reads。

然后利用序列综合分析软件 jellyfish3.3.1 绘制 k-mer 分布图，并参照公式 M=N*(L-K+1)/L，计算实际测序深度，其中 N 代表实际测序深度，M 代表 k-mer 曲线的峰值，L 代表 reads 的长度，k 代表字符串的碱基数量。最后，把全部序列长度除以实际测序深度 N，预测基因组大小。

将 Illumina 数据利用软件 Velvet[21]进行拼接，组装获得基因组草图。然后将拼接组装后的 contig，与 NCBI 中现有的 1,002 条雨生红球藻 EST 序列进行比对，分析基因组草图覆盖率。拼装出的基因组草图，使用 Augustus[22]进行基因预测。将预测基因的编码蛋白，通过 blastp 分别与 NCBI 的非冗余蛋白数据库(nr)、KEGG (Kyoto Encyclopedia of Genes and Genomes)和 UNIPORT (Universal Protein)数据库比对，进行功能注释，比对参数设为：E-value <1e^{-5}。依据注释结果，绘制 KEGG 代谢通路图；利用 CDD 数据库[23]和 RPS-Blast 进行 KOG 分类，进一步进行功能注释。

2. 结果与分析

2.1 数据预处理和拼接

测序共得到 38,189,673 对序列，测序长度是 2*150 bp，总碱基数是 11.45 G，过滤得到高质量基因组序列 37,185,329 对，碱基数约 10.55 G。通过对原始数据碱基组成分布和原始数据碱基质量分布进行分析，发现该文库每个位置上，A、C、G、T 在开始有所波动，后面会趋于稳定。一般情况下 A 与 T 相等，C 与 G 相等，各碱基所占百分比会因物种差异而不同。该文库碱基分布均匀，N 的比例非常低，碱基质量良好，可用于后续分析。

2.2 基因组大小预测与杂合度分析

k-mer 是指将 reads 分成包含 k 个碱基的字符串，一般长度为 L 的 reads 可以分成 L-k+1 个 k-mers。本研究中每条 150 bp 的序列可以拆成 134 个 17 bp 长的字符串，得到测序 reads 所有 17-mer 片段，然后统计各 17-mer 及其出现的次数（k-mer frequnce）。以出现次数（depth）为横坐标，以出现多次的片段总数为纵坐标作图，那么曲线的峰值 M，就接近实际测序深度 N。参照公式：M=N*(L-K+1)/L，然后把全部序列长度除以实际测序深度 N，获得预测的基因组大小。然后利用序列综合分析软件 jellyfish3.3.1，统计出 10.55 Gb clean reads 中的 17-mer，并作图（图 1），得出 M=17，实际测序深度 N=19。一般认为只出现 1 次或几次的 17-mer 是测序错误，根据图 3，最左侧的谷底 17-mer 对应深度为 8，因此去除出现频率< 8 的全部 17-mer，则有效数据为全部数据的 98.5%。所以雨生红球藻的基因组大小估计为 G=10,556,751,790*0.985/19=547 Mb。同时，由图 3 可以看出只有一个峰，表示雨生红球藻为纯合或单倍体。

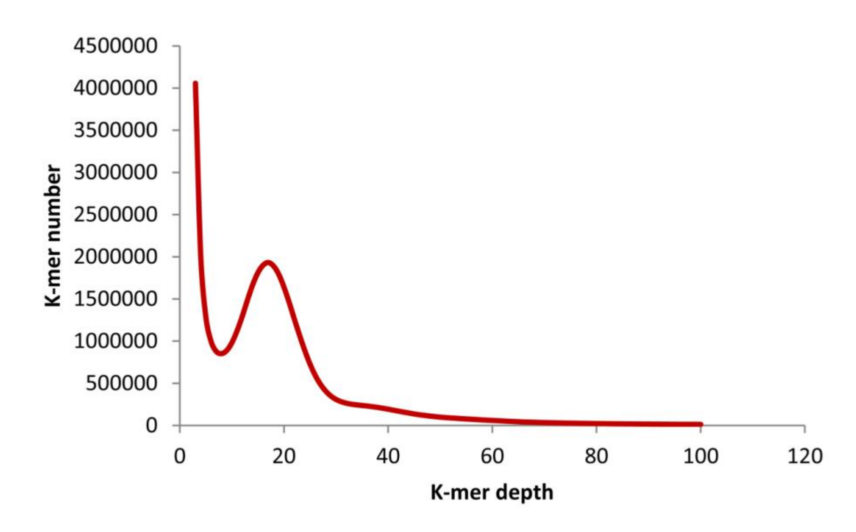


图 1　雨生红球藻基因组序列的 17-mer 分析

Fig. 1 17-mer analysis of genome contigs in *Haematococcus pluvialis*

2. 3 序列组装和基因组草图覆盖率评价

表 1　不同 k-mer 下的基因组组装结果

Table 1 the result of genome assemble with different k-mer value

k-mer	Contigs Number	Size (bp)	Ave Length (bp)	Reads Usage
35	113061	66423650	587	24.60%
45	118556	87506031	738	30.60%
55	63142	96909627	1534	34.80%
65	58115	101883098	1753	39.10%
75	56423	104818003	1857	42.00%
85	58995	106127310	1798	43.60%
95	66985	105643283	1577	44.40%

利用生物信息学软件 Velvet 对 Illumina 数据进行拼接，得到了不同 k-mer 值下基因组组装结果，如表 1 可见，不同的 k-mer 值对 contig 的数量和总长度影响均较显著。一般选择拼装结果，采取的原则是 contig 尽量少，contig 的长度尽量大。所以，在一定的范围内，就是选择 contig 平均长度最大（contig 的长度除以 contig 数目，获得 contig 平均长度）的拼装结果。故本研究中采用 k-mer 值为 75 的情况下的拼装结果，即获得 56,423 条 contig，总长度为 104,818,003 bp，contig 的平均长度为 1,857 bp，平均 GC 含量为 59.2 %。由于有效覆盖率只有 17 倍，因此只能组装出部分基因组序列。利用 NCBI 中现有的 1,002 条 EST 序列与测序获得的 contig 比对，结果显示 910 条序列可以比对到测序拼接的基因组，占总数的 90.82 % ；495 条 EST 90%以上序列被单个 contig 覆盖，比例为 49.40；782 条 EST 50%以上序列可以被单个 contig 覆盖，比例为 78.04（表 2）。将目前 NCBI 中现有的 EST 序列与我们预测的基因进行比对，有 412 条 EST 序列可以匹配，占总数的 41.12 %。雨生红球藻 EST 序列覆盖度分析表明本草图覆盖了雨生红球藻基因编码区的 90 %左右，表明获得了雨生红球藻低覆盖度基因组草图。

表 2 以已知雨生红球藻 EST 序列评价基因组草图完整性

Table 2 Assessment the sequence coverage of *H. pluvialis* draft genome using known ESTs

长度覆盖率（%）	匹配的 EST 累积数目	累积比例（%）
100	264	26.347
90	495	49.401
80	576	57.485
70	664	66.267
60	727	72.554
50	782	78.043
40	821	81.936
30	849	84.73
20	877	87.524
10	910	90.818

利用测序获得的 contig 与 NCBI 中现有的藻类基因组序列进行比对，如表 3 所示，共有 4964 条 contig 可以比对到雨生红球藻基因序列，匹配的 contig 总长度为 9534797 bp，比对长度为 696150；其次，匹配到莱茵衣藻（*Chlamydomonas reinhardtii*）的 contig 数量为 1977 条，匹配的 contig 总长度为 4068934 bp，比对长度为 246989 bp。莱茵衣藻是绿藻门模式生物，与雨生红球藻同属绿藻门，团藻属，匹配同源性较高符合预期。

表 3 以已知 NCBI 藻类基因组序列评价基因组草图完整性

Table 3 Assessment the sequence coverage of *H. pluvialis* draft genome using known algal genome sequences updated from NCBI

匹配基因组	匹配的 contig 数量	匹配的 contig 总长度（bp）	比对长度 (bp)
雨生红球藻 (*Haematococcus pluvialis*)	4964	9534797	696150
莱茵衣藻 (*Chlamydomonas reinhardtii*)	1977	4068934	246989
团藻 (*Volvox carteri* f. nagariensis)	401	800015	45726
多变小球藻 (*Chlorella variabilis*)	347	747501	38375
胶球藻 C-169 (*Coccomyxa subellipsoidea C-169*)	178	352938	20464
原壳小球藻 (*Auxenochlorella protothecoides*)	165	316284	16626

2.4 基因预测与功能注释

用拼装出的基因组，使用 Augustus 进行基因预测，共得到 11,059 个预测基因，平均基因长度为 1,711 bp，平均 CDS 长度为 681 bp；平均每个基因包含 3.2 个外显子，外显子平均长度为 353 bp。将预测基因的编码蛋白，通过 blastp 分别与 NCBI 的非冗余蛋白数据库（nr）、KEGG (Kyoto Encyclopedia of Genes and Genomes) 和 UNIPORT (Universal Protein)数据库比对，进行功能注释，结果共有 6,890 个蛋白具有明确的生物学功能，3,117 个蛋白具有 KEGG 的 ortholog，注释率为 62.30 %。利用预测的 6,890 个蛋白与 NCBI 中现有的藻类蛋白序列进行比对，如表 4 所示，预测基因所匹配的同源性最高的蛋白所属的物种分别是团藻（*Volvox carteri* f. nagariensis）、莱茵衣藻（*Chlamydomonas reinhardtii*）、单针藻（*Monoraphidium neglectum*）、胶球藻 C-169（*Coccomyxa subellipsoidea* C-169）、多变小球藻（*Chlorella variabilis*）、原壳小球藻（*Auxenochlorella protothecoides*）、雨生红球藻（*Haematococcus pluvialis*）和盐生杜氏藻（*Dunaliella salina*）。其中，E-value 为 10^{-3} 的情况下，预测蛋白与团藻蛋白序列匹配的数目为 2148 条，占比为 24.06%。

根据 KEGG 注释结果生成代谢通路图，共获得 230 个代谢通路图，具有完整的糖酵解、三羧酸循环、磷酸戊糖途径、嘌呤和嘧啶合成等基本代谢通路，说明拼装得到的序列中包含了基本的必需基因信息。其中注释率较高的代谢通路主要包括：碳水化合物代谢（Carbohydrate metabolism），氨基酸代谢（Amino acid

metabolism），能量代谢（Energy metabolism），蛋白质翻译（Translation），辅因子和维生素代谢（Metabolism of cofactors and vitamins），蛋白质折叠、分选与降解（Folding, sorting and degradation），膜转运（Membrane transport），信号转导（Signal transduction）， 脂代谢（Lipid metabolism）等通路（图 2）。

表 4 以已知藻类蛋白序列评价基因组草图完整性

Table 4 Assessment the sequence coverage of *H. pluvialis* draft genome using known algal protein sequences

最佳匹配物种	**E-value=10^{-3}**		**E-value=10^{-10}**		**E-value=10^{-20}**	
	数目	**比例**	**数目**	**比例**	**数目**	**比例**
团藻 *Volvox carteri* f. nagariensis	2148	24.06%	1900	23.57%	1511	21.92%
莱茵衣藻 *Chlamydomonas reinhardtii*	1820	20.39%	1618	20.07%	1315	19.08%
单针藻 *Monoraphidium neglectum*	516	7.49%	433	6.28%	313	4.54%
胶球藻 *Coccomyxa subellipsoidea* C-169	262	2.93%	225	2.79%	163	2.37%
多变小球藻 *Chlorella variabilis*	172	1.93%	143	1.77%	90	1.31%
原壳小球藻 *Auxenochlorella protothecoides*	72	0.81%	56	0.69%	38	0.55%
雨生红球藻 *Haematococcus pluvialis*	68	0.76%	68	0.84%	66	0.96%
盐生杜氏藻 *Dunaliella salina*	41	0.45%	39	0.48%	37	0.54%
其他	1260	14.11%	1011	12.54%	796	11.55%

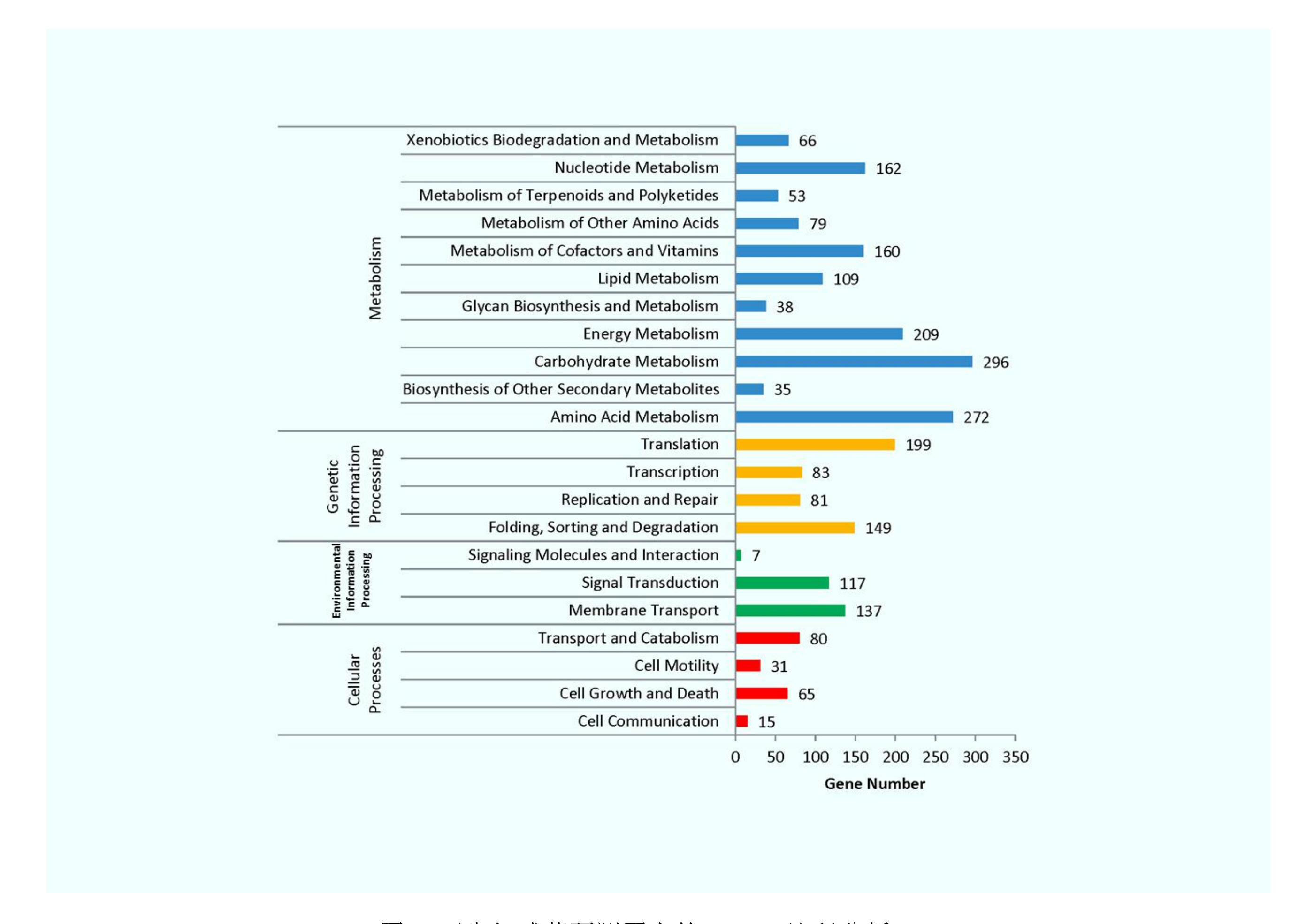


图 2 雨生红球藻预测蛋白的 KEGG 注释分析

Fig. 2 KEGG analysis for predicted proteins of *Haematococcus pluvialis*

利用 CDD 数据库和 RPS-Blast 进行 KOG 分类，结果共有 5,233 个蛋白被 KOG 分类注释，按照功能不同可分成 26 个类群，主要包括一般功能预测（General function prediction only），翻译后修饰、蛋白折叠及操纵子（Posttranslational modification, protein turnover, chaperones），信号转导机制（Signal transduction mechanisms），氨基酸转运与代谢（Amino acid transport and metabolism），蛋白翻译、核糖体结构与合成（Translation, ribosomal structure and biogenesis），碳水化合物转运与代谢（Carbohydrate transport and metabolism）（图 3）。在不同 E-value 下的 KOG 比对结果如表 5 所示。

表 5 预测蛋白在不同 E-value 下的 KOG 分类

Table 5 Assessment the KOG analysis for predicted proteins of *Haematococcus pluvialis* under different E-values

KOG 分类	1.00E-03	1.00E-05	1.00E-10	1.00E-15	1.00E-20	1.00E-25	1.00E-30
Amino acid transport and metabolism	363	343	310	281	245	221	192
Carbohydrate transport and metabolism	334	314	256	213	179	149	129
Cell cycle control, cell division, chromosome partitioning	112	97	70	59	46	35	27
Cell motility	1	1	1	0	0	0	0
Cell wall/membrane/envelope biogenesis	57	50	34	29	17	14	14
Chromatin structure and dynamics	73	67	53	48	43	34	23
Coenzyme transport and metabolism	124	115	101	87	81	71	56
Cytoskeleton	213	179	149	124	107	91	74
Defense mechanisms	21	20	16	6	2	2	2
Energy production and conversion	286	269	243	207	184	162	141
Extracellular structures	49	21	5	2	2	2	2
Function unknown	177	146	103	70	50	38	25
General function prediction	482	410	293	236	196	167	132
Inorganic ion transport and metabolism	231	201	161	126	100	84	64
Intracellular trafficking, secretion, and vesicular transport	220	191	170	139	111	91	79
Lipid transport and metabolism	209	193	159	134	111	88	76
Nuclear structure	16	14	10	7	5	4	3
Nucleotide transport and metabolism	148	141	130	110	97	83	66
Posttranslational modification, protein turnover, chaperones	470	434	364	312	259	218	185
Replication, recombination and repair	152	138	114	100	88	72	58
RNA processing and modification	278	261	229	190	158	130	110
Secondary metabolites biosynthesis, transport and catabolism	227	216	189	156	133	96	71
Signal transduction mechanisms	391	325	253	200	157	117	88
Transcription	240	195	142	109	91	69	57
Translation, ribosomal structure and biogenesis	359	346	308	271	226	198	167
总计	**5233**	**4687**	**3863**	**3216**	**2688**	**2236**	**1841**

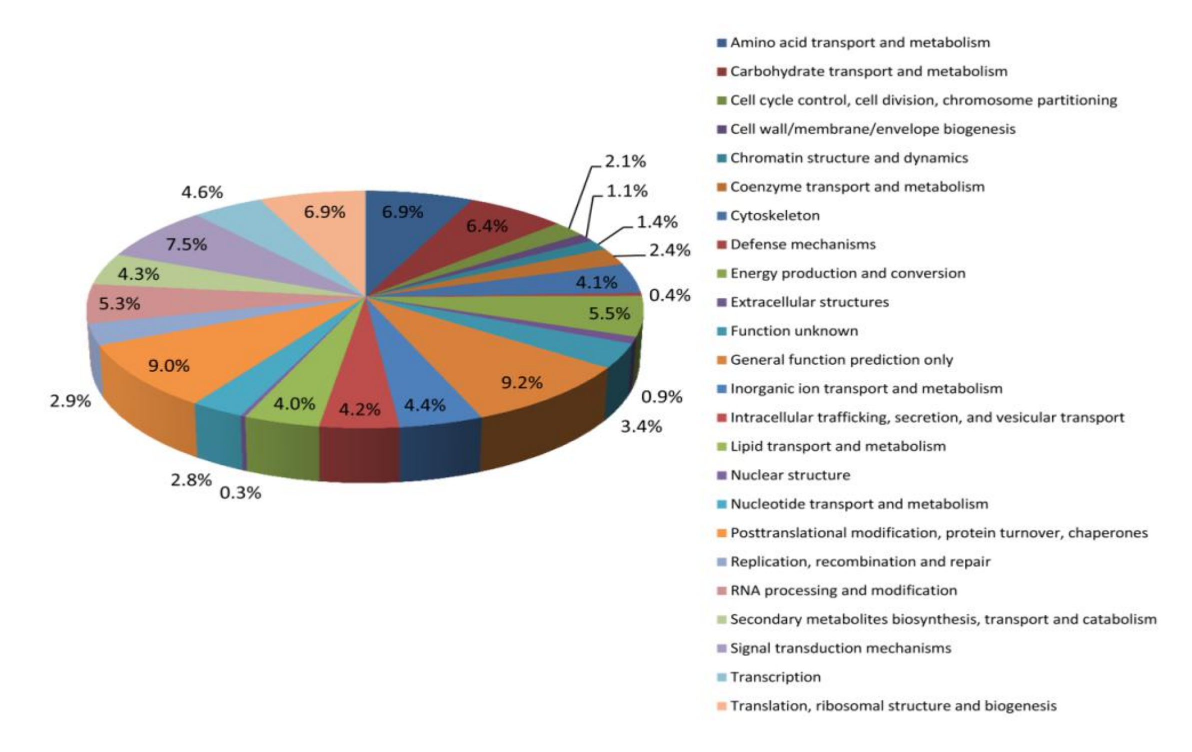


图 3 雨生红球藻预测蛋白的 KOG 分类

Fig. 3 KOG analysis for predicted proteins of *Haematococcus pluvialis*

3. 讨论

真核藻类起源于内共生事件，种类繁多，进化地位特殊、过程复杂。内共生学说认为，原核生物蓝藻经过初级内共生以后，形成绿藻、红藻和灰藻；在绿藻和红藻的基础上，经过次级内共生形成了其他微藻，例如隐藻（*Cryptophycean*）。其中，绿藻形态多样，光合色素系统与高等植物相似，既含有叶绿素 a，也含有叶绿素 b，其进化地位虽介于高等植物和原核蓝藻之间，但又不可简单的看成是介于微生物和高等植物之间的过渡类群，长期以来也一直是植物学界和藻类学界研究的热点。开展雨生红球藻基因组测序研究，对于解析绿藻门演化趋向、真核藻类起源与演化等方面都具有极其重要的意义。

雨生红球藻分布广泛，耐受性强，能够适应缺氮、高光、低氧、高盐等特殊逆境，并合成虾青素。复杂的逆境孕育了其精细的信号转导系统和特殊的次生代谢系统，是研究生物逆境响应的理想材料。开展雨生红球藻的全基因组测序研究，将为建立雨生红球藻的基因组、转录组、代谢物组等系统生物学模式研究体系，深入地考察雨生红球藻能量存储和转化、逆境胁迫响应机理等一系列复杂性状的网络调控机制，从而建立雨生红球藻的基因调控网络的功能基因组学研究模型，为直接指引通过代谢工程手段和合成生物学的思路改造和构建高质量的藻株提供重要参考。

综上，无论是从研究真核藻类起源与进化的角度，还是推动雨生红球藻产业发展的层面，雨生红球藻的全基因组测序都亟待尽快被完成。本研究利用第二代高通量测序技术开展了雨生红球藻低覆盖度全基因组测序，并预测了雨生红球藻基因组大小等基因组基本特征，为绘制高质量雨生红球藻基因组精细图作了准备，但与近年来报道的高质量绿藻基因组精细图谱比较仍有差距。例如，Roth 等（2017）采用三代 Pacbio 和二代基因组测序技术相结合，完成了佐夫色绿藻染色体水平的基因组组装，共组装出 19 条染色体，发现编码基因 1.5 万多个。然后，采用转录组学技术和通过测定藻株类胡萝卜素含量，系统解读了雨生红球藻野生株和突变株响应不同光强胁迫条件时的虾青素积累机制。该研究不仅绘制了高质量的佐夫色绿藻基因组，并详细解读了有关该藻的重要科学问题，如不同光强胁迫条件下虾青素积累机制，为后续利用代谢工程等手段构建优良藻株提供了理论依据，值得后续开展雨生红球藻全基因组精细图谱绘制时参考。

开展雨生红球藻全基因组测序及精细图谱的绘制，需要注意以下几方面问题：（1）建立雨生红球藻无菌化培养体系。雨生红球藻常常伴有共生菌，且利用 EG:JM 等有机培养基培养时容易被细菌污染，开展全基因组测序研究需要特别注意无菌化处理，以避免测序获得的基因组拼接困难等问题。郑凌凌等（2017）研究发现，对于雨生红球藻 FACHB-712 藻株可选择青霉素、庆大霉素、卡那霉素两种或 3 种依次加入，实现藻株除菌操作[25]。本研究在开展全基因组测序前，特别注意了染菌问题，先后通过反复纯化、显微镜镜检，16S rRNA 测序等手段，确保测序样品未受到细菌污染。（2）建议将二代测序技术与三代测序技术结合运用。基于本研究预测发现，雨生红球藻基因组大约 500M 左右，且应该含有多条线性染色体，建议采用最新的第三代测序系统 PacBio SMRT 测序。该测序系统可以获得长达 10 Kb 的 reads，拼装效果较好。同时，辅以 Illumina Hiseq2500 测序，可以校正 PacBio 测序中的单碱基 indel 的错误，以便提高拼装结果的准确性。

参考文献：

[1] 胡鸿钧，魏印心. 中国淡水藻志-系统、分类及生态[M]. 北京：科学出版社, 2006.

Hu H J, Wei Y X. The freshwater algae of China- Systematics, Taxonomy and Ecology[M]. Beijing: Science Press, 2006 (in Chinese).

[2] 刘建国，殷明焱，张京浦, *et al*. 雨生红球藻的细胞周期初探. 海洋与湖沼, 2000, 02: 145-150.

Liu J G, Yin M Y, Zhang J P, *et al*. Statues of cell cycle in *Haematococcus pluvialis*[J]. Oceanology et Limnologia Sinica, 2000, (02): 145-150 (in Chinese).

[3] Kobayashi M, Kurimura Y, Kakizono T, *et al*. Morphological changes in the life cycle of the green alga *Haematococcus pluvialis*[J]. Journal of Fermentation and Bioengineering, 1997, 84(1): 94-97.

[4] Lorenz R T, Cysewski G R. Commercial potential for *Haematococcus* microalgae as a natural source of staxanthin[J]. Trends in Biotechnology, 2000, 18(4):160-167.

[5] Borowitzka M A, High-value products from microalgae-their development and commercialisation[J]. Journal of

Applied Phycology, 2013, 25(3): 743-756.

[6] Chen J, Wang Y, Benemann J R, *et al*. Microalgal industry in China: challenges and prospects[J]. Journal of Applied Phycology, 2016, 28(2): 715-725.

[7] Cui H L, Yu X N, Wang Y, *et al*. Evolutionary origins, molecular cloning and expression of carotenoid hydroxylases in eukaryotic photosynthetic algae[J]. BMC Genomics, 2013, 14: 457.

[8] Han D X, Li Y T, Hu Q. Astaxanthin in microalgae: pathways, functions and biotechnological implications[J]. Algae, 2013, 28(2): 131-147.

[9] Lu Y D, Jiang P, Liu S F, *et al*. Methyl jasmonate- or gibberellins A(3)-induced astaxanthin accumulation is associated with up-regulation of transcription of beta-carotene ketolase genes (bkts) in microalga *Haematococcus pluvialis*[J]. Bioresource Technology, 2010, 101(16): 6468-6474.

[10]Gao Z Q, Li Y, Wu G X, *et al*. Transcriptome analysis in *Haematococcus pluvialis*: Astaxanthin induction by salicylic acid (SA) and jasmonic acid (JA)[J]. Plos One, 2015, 10(10): e0140609.

[11]Li K, Cheng J, Lu H X, *et al*. Transcriptome-based analysis on carbon metabolism of *Haematococcus pluvialis* mutant under 15% CO2[J]. Bioresource Technology, 2017, 233: 313-321.

[12]Cheng J, Li K, Zhu Y X, *et al*. Transcriptome sequencing and metabolic pathways of astaxanthin accumulated in *Haematococcus pluvialis* mutant under 15% CO2[J]. Bioresource Technology, 2017, 228: 99-105.

[13]Chen G Q, Wang B B, Han D X, *et al*. Molecular mechanisms of the coordination between astaxanthin and fatty acid biosynthesis in *Haematococcus pluvialis* (Chlorophyceae)[J]. Plant Journal, 2015, 81(1): 95-107.

[14]Gwak Y, Hwang Y S, Wang B B, *et al*. Comparative analyses of lipidomes and transcriptomes reveal a concerted action of multiple defensive systems against photooxidative stress in *Haematococcus pluvialis*[J]. Journal of Experimental Botany, 2014, 65(15): 4317-4334.

[15]Wang S B, Chen F, Sommerfeld M, *et al*. Proteomic analysis of molecular response to oxidative stress by the green alga *Haematococcus pluvialis* (Chlorophyceae)[J]. Planta, 2004, 220(1): 17-29.

[16]Kim J D, Lee W S, Kim B, *et al*. Proteomic analysis of protein expression patterns associated with astaxanthin accumulation by green alga *Haematococcus pluvialis* (Chlorophyceae) under high light stress[J]. Journal of Microbiology and Biotechnology, 2006, 16(8): 1222-1228.

[17]Gao Z Q, Miao X X, Zhang X W, *et al*. Comparative fatty acid transcriptomic test and iTRAQ-based proteomic analysis in *Haematococcus pluvialis* upon salicylic acid (SA) and jasmonic acid (JA) inductions[J]. Algal Research-Biomass Biofuels and Bioproducts, 2016, 17: 277-284.

[18]Su Y X, Wang J X, Shi M L, *et al*. Metabolomic and network analysis of astaxanthin-producing *Haematococcus pluvialis* under various stress conditions[J]. Bioresource Technology, 2014, 170: 522-529.

[19]Lv H X, Xia F, Liu M, *et al*. Metabolomic profiling of the astaxanthin accumulation process induced by high light in *Haematococcus pluvialis*[J]. Algal Research-Biomass Biofuels and Bioproducts, 2016, 20: 35-43.

[20]Boussiba, S. Carotenogenesis in the green alga *Haematococcus pluvialis*: Cellular physiology and stress response[J]. Physiologia Plantarum, 2000, 108(2): 111-117.

[21]Zerbino D R, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs[J]. Genome Research, 2008, 8(5): 821-829.

[22]Stanke M, Schoffmann O, Morgenstern B, *et al*. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources[J]. BMC Bioinformatics, 2006, 7: 62.

[23]Marchler-Bauer A, Bo Y, Han L, *et al*. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures[J]. Nucleic Acids Research, 2017, 45(D1): D200-D203.

[24]Roth M S, Cokus S J, Gallaher S D, *et al*. Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, E4296–E4305.

[25]郑凌凌，张琪，李天丽, *et al*. 雨生红球藻无菌化处理及其对生长和生理的影响[J]. 福建师范大学学报(自然科学版), 2017, (01): 44-50.

Zheng L L, Zhang Q，Li T L，*et al*. Axenation of *Haematococcus pluvialis* and the effects of axenic cultivation on the growth and physiology of the strain[J]. Journal of Fujian Normal University (Natural Science Edition), 2017, 33(1): 44-50 (in Chinese).